Plant Disease Detection using Artificial Intelligence and Machine Learning

A PROJECT REPORT

Submitted in partial fulfillment for the Degree of

Bachelor of Technology under the

School of Computing

Submitted by

Register No	Names of Students
AM.EN.U4AIE21122	Bollam Shiva Shankara Varaprasad
AM.EN.U4AIE21140	Lekha Sathvik Devabathini
AM.EN.U4AIE21157	Satvik Choulapally
AM.EN.U4AIE21169	Yasin Syed

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

Amritapuri Campus (INDIA)

December 2024

Department of Computer Science and Engineering

AMRITA VISHWA VIDYAPEETHAM

AMRITAPURI CAMPUS

BONAFIDE CERTIFICATE

This is to certify that this is a bonafide record of the project presented by the students whose names are given below during Project Phase 1 in partial fulfilment of the requirements of the degree of Bachelor of Technology in Computer Science and Engineering.

Roll No	Names of Students
AM.EN.U4AIE21122	Bollam Shiva Shankara Varaprasad
AM.EN.U4AIE21140	Lekha Sathvik Devabathini
AM.EN.U4AIE21157	Satvik Choulapally
AM.EN.U4AIE21169	Yasin Syed

Dr. Gopakumar G.
(Project Guide)

<Coordinator name here>
(Project Coordinator)
Designation

<Reviewer's Name>
(Reviewer)
Designation

Contents

Acknowledgements

We want to express our sincere gratitude to those who helped and contributed to completing the project "**Plant Disease Detection using Artificial Intelligence and Machine Learning**". We are grateful for everyone's support and guidance to enable us to research and engineer to contribute to the sector of agriculture.

We express our sincere gratitude to our guide Dr. Gopakumar G., for guiding, helping and supporting us in every step of this project. Dr. Gopakumar G.'s guidance and expertise have been invaluable throughout the research and development done for this project. It would have been impossible to complete this project without his care and support

We want to thank our advisor and organizer **Raji** for keeping us organized and guiding us to complete the project on time and providing us with a supportive environment to complete our project. Her contribution to completing this project is priceless.

We are grateful for the support of Dr. Divya R for guiding us to effectively complete our project and see that we don't get derailed in our research.

We want to express our gratitude to our teachers and friends for their helpful insights and constructive feedback. Their insights and opinions have been invaluable for enriching and improving our project. We are grateful for their time and effort in participating in some of the most valuable discussions to guide and improve the project and help in its successful completion.

Bollam Shiva Shankara Varaprasad
Lekha Sathvik Devabathini
Satvik Choulapally
Yasin Syed

List of Figures

List of Tables

Abstract

India is an agricultural country, as two-thirds of its population is engaged in agricultural activities, the development in this sector is very important for the country both economically and socially. The losses in agriculture due to pests and diseases are estimated to be approximately 20-25%. Despite the significance of the agricultural sector for the country, the application and implementation of Artificial intelligence and machine learning have been limited in the past five years. We must introduce cutting-edge technologies in Artificial intelligence to the agricultural sector to improve the quality of life of farmers and other people in the sector. This project introduces a solution to diagnose the common pests and diseases in plants using Artificial intelligence and Machine learning techniques. The study focuses on implementing a robust solution to classify different plant diseases based on the images mainly focused on the leaves. The project proposes a Vision Transformer(ViT)-based solution with an accuracy of 99 per cent on the test dataset for diagnosis of fourteen different plant diseases across three different crops. The lack of introduction of these technologies can also be interpreted as the consequence absence of good, clean and homogeneous data for training different models. To help the future development of similar solutions the project summarizes the current state of the various famous datasets for usage in production-level solutions development. As the solutions developed are highly dependent on the availability and quality of existing datasets, we propose a clean and efficient way to collect and organize more data for further development. The project also implements and tests how these solutions can be efficiently deployed and made accessible to everyone. The project starts out with an essential focus on the differentiation of the chilli plants that are infected by the leaf curl virus form the healthy ones to eventually generalize the problem of plant disease diagnosis. Large Language Models(LLM's) are the current trend in making data more accessible to everyone. The project also engineers a solution to introduce the LLM's to make the user more knowledgeable about the disease at hand and also guide him in the process of minimizing the loss due to the pest or disease.

Chapter 1

Introduction

1.1 Background

The sector of agriculture is one of the most important sectors for the Indian economy. The agricultural sector contributes about 20-25% to the GDP of India. There has been a steady decline in the contribution of the agricultural sector to the Indian economy. This can be interpreted as a result of lack of research and development in this area. Especially in India agricultural sector has the least amount of growth and development compared to the other sectors which are actively implementing and adopting the cutting-edge technologies for rapid development. The introduction of Artificial intelligence seem to be the slowest in the agricultural sector compared to its explosive growth in other industries. This leaves the agricultural sector with a huge gap to incorporate and adopt the cutting-edge Artificial Intelligence and Machine Learning techniques for further improvement. This motivates us to research and develop AI and ML based solutions to solve the common problems in the agricultural sector. Even though there has been a considerable amount of research to introduce Artificial intelligence techniques into agriculture this project keeps the accessibility of the research and development as its top priority.

Through the previous economic surveys the loss of income due to plat diseases and pests is about 20-25%, which is a significant figure over the total economy of the agriculture. The areas of Artificial Intelligence and Machine Learning have seen an explosive development in the past few years. With the latest advancements in computer vision and natural language processing it is possible to device solutions to minimize the costs due to plant diseases and pests in agriculture. Most of the research done to incorporate these techniques never was really adopted and applied in the real world conditions. All these circumstances and conditions form the basis for this project. As most of the solutions that can be incorporated from the areas of Artificial Intelligence and Machine Learning are heavily dependant on the amount and quality of data, it is essential to improve the data resources and repositories to enable further research and development. It is very essential to analyze the available resources and weigh their pros, cons and biases to build robust production ready systems that are ready for use

in fields. The study hypothesizes that it is possible to develop a solution accessible through widely available hardware to detect diseases in plants early and provide suitable solutions to counter the progression of that disease.

1.2 Problem Addressed

The project aims to introduce various Artificial Intelligence and Machine Learning techniques to the field of agriculture to minimize the losses due to pests and diseases. The study focuses on the detection and the possibility of early detection of various diseases in plants by using AI (Artificial Intelligence) and ML (Machine Learning). This study focuses on assessing the feasibility of making the developed solution available to regular people with minimal resources.

The study explores the feasibility of various machine learning techniques such as Support Vector machines, DNNs (Deep Neural Networks, RNNs (Recurrent Neural Networks) and Transformers for solving the task.

the problem is to develop a solution for early detection and detection of diseases in plants with minimal requirements that are accessible by everyone by exploring a wide range of machine learning and deep learning techniques. Focus on the feasibility of the developed solutions to be applied easily by normal people.

The recent advancements in natural language processing through the use of Large Language Models(LLMs) opened new ways to efficiently manage knowledge and data. The project deals with the problem of guiding the user further based on a predefined knowledge to minimize the loss due to the disease.

The solutions built using the AI and ML techniques are known to be computationally expensive to deploy and scale. The project aims to optimize the deployment of the proposed solution over limited resources. The deployment architecture focuses of effective horizontal scaling for maximum accessibility to the masses. It is important that the solutions that are developed in the research are reaching the people to make some difference.

The proposed solution should be able to pass the test that the internal testing team does to prove its viability. The developed solution should be satisfactory to the initial users to continue for the further development. The model should show reliable results on the real world data that it is exposed to during the testing in fields. The current testing focus is majorly on the Chilli Leaf Curl virus and the ability of the model to differentiate the infected plant from the healthy ones.

As the solutions that are currently being explored and the ones that are going to be devel-

oped in the future are heavily dependant on the quality and availability of data for training, we aim to provide a critical review on the state of various datasets with emphasis on pros, cons and biases observed in each of them. The problems that arise when combining multiple datasets to make a bigger dataset should also be effectively explained for the future research and development.

The study aims to provide well-documented research on the performance of a wide range of AI and ML techniques to detect diseases in plants. The project also focuses on the development of a mobile app that can be used to implement these modes on edge with a focus on accessibility on consumer hardware.

1.3 Motivation

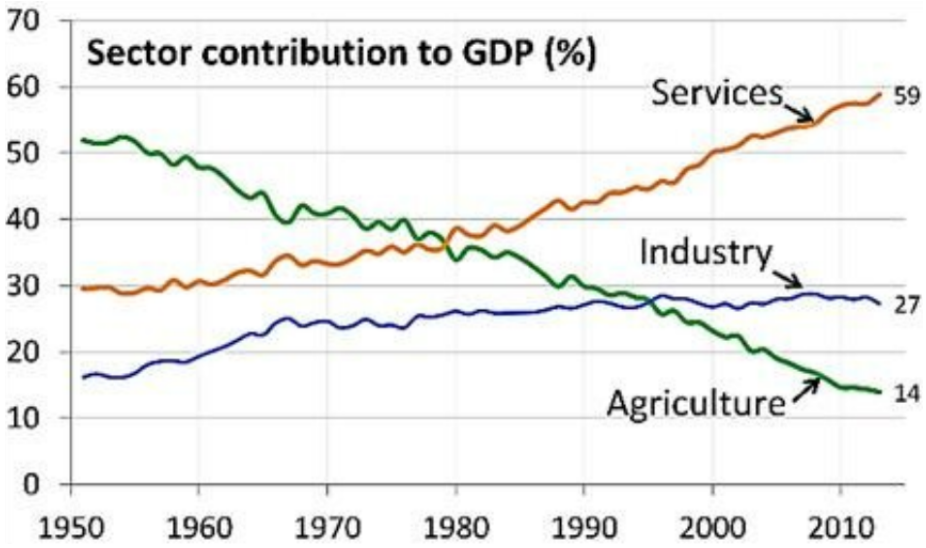

Figure 1.1: Decline in contribution of agriculture to India's GDP

The occupations based on agriculture are becoming harder to sustain due to the environmental changes and unexpected changes in the climate. The field of agriculture has seen less growth and development compared to the other sectors mainly through the explosive development on Artificial Intelligence and Machine Learning techniques. The introduction of these latest techniques have been slow in the areas of agriculture compared to other sectors. We aim to contribute to the sector of agriculture by exploring some of the possible solutions to diagnose plat diseases and pests to minimize the losses to the farmers.

We believe that the research gets value when solutions are engineered based on the research. This project aims to analyze the complexities and optimizations in the deployment of the developed solutions. We are motivated by the fact that the knowledge is only useful when it is applied at the right place and the right time. The project aims to enable the users to get the required knowledge about the plant diseases and pests in the right time.

It is important to facilitate the future research and development through a critical review of the datasets that are used in the project based on their quality and document the challenges faced during the usage of these datasets.

The project progresses on the motivation to introduce the latest Artificial Intelligence and Machine Learning techniques to the field of agriculture to improve the overall conditions of the sector.

1.4 Scope of the Project

The study starts small with a focus on the leaf curl virus and aims to generalize the solution to various common diseases in plants. The project focuses on the edge deployment of the solution on consumer hardware and these aspects set apart the current project from others.

The study aims to act as a first step in building sustainable and accessible AI and ML solutions to contribute to the agricultural sector. The study currently proposes a prototype to diagnose plant diseases and give some basic knowledge about the detected disease to the user. The testing scope of the project includes the testing of the built prototype in field for detection of Leaf Curl Virus in Chilli plants.

The project does not aim to provide a completely accurate application to diagnose the diseases in plants but to act as a proof of concept and guide to build more advanced and sophisticated solutions to contribute to the field of agriculture. The project introduces efficiently scalable deployment architecture for computationally expensive Artificial Intelligence and Machine Learning solutions to make the cutting edge research accessible to the people.

Chapter 2

Literature Review/ Existing System

2.1 Existing System Study

There have been a significant amount of research to introduce Artificial Intelligence and Machine Learning technologies to the field of agriculture to minimize the losses due to diseases and pests. These researches and solutions help us build better solutions in the current project. An extensive study is done to explore and analyze the existing systems and solutions that are trying to solve similar problems.

In the paper **Construction of deep learning-based disease detection model in plants** Minah Jung et al. proposed a Deep Learning based framework for the construction of disease detection models in plants. The solution is designed to work in three steps namely: Crop Classification, Disease Detection and Disease Classification [1]. The segregation of responsibility and labour in the proposed solution can be leveraged to efficiently scale the built solution. Also as the division on labour is clear and strict the retraining cost for the solution can be minimized by limiting the changes to the concerned model. Even though the solution developed in the paper have inconsistent performance across different plant species, the modularity of the architecture allows us to correct what went wrong without disturbing what went right. Even though the proposed solution can be horizontally scaled on demand due to it's well defined modularity, an ill deployed solution can be computationally expensive and can be heavy on computation due to large chain of action. This introduces the cost of latency in return for the ease of systematic and efficient training and deployment.

In the paper **Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images** , Vibhav et al. proposed a Dense Neural Network architecture for plant disease classification. The proposed solution is highly accurate on the used dataset with an impressive response time of 16ms per image for classification. Even though the proposed solution's performance is inconsistent across different classes, the paper sheds light on how a centralized monolithic model performs in the disease detection tasks. In contrast to the Minah Jung et al's work the current solution may not have a good division of

responsibility and retraining efficiency, the solution boasts a significantly better performance in deployment. The paper does not provide any information about the steps taken to clean or balance the dataset for training.

Do a literature survey and identify five most suitable base papers that align with your project. Select the top journal/conference paper which are recent in the area, which has addressed this problem. For each of those papers, write a paragraph that subsumes, Contributions of the paper: (A short paragraph of less than 50 words) , Limitations of the paper : (A short paragraph of less than 50 words) , Open Problems/Future work possible: (Enumerate the list)

Review existing products, tools, or technologies solving similar problems. Include market comparisons or feature analysis.

2.2 Research Gap/ Scope for improvement and innovation

Limitations/ problems not addressed in existing systems Point out the specific shortcomings or challenges that persist in existing solutions.Identify areas or research questions that have not been addressed

1. Which are all the gaps that you have addressed now in the Phase I
2. Gaps which you plan to address in phase II(If the project is to be continued)

2.3 Problem Statement and Contributions

1. Problem Statement: Clearly articulate the specific challenge you aim to solve. Ensure it flows naturally from the gaps identified in the previous slide. Use precise, concise language to define the problem scope.
2. Research Contributions Present 2-4 clear, measurable research contribution of your work. For Research Projects: Focus on advancing knowledge, developing methodologies, or validating hypotheses. For Product-Based Projects: Focus on delivering a functional, innovative solution with specific features or improvements.

Chapter 3

Proposed Work

3.1 Proped Work

Provide an overview of the proposed work, connecting it to the research gaps identified in the Literature Study chapter. Briefly restate the problem. Highlight how your proposed work addresses the gaps or challenges. Emphasize the novelty or uniqueness of your approach.

3.1.1 Objectives of the Proposed Work

Clearly articulate the specific objectives of the proposed research or system.

Define measurable goals. Include both primary and secondary objectives.

3.2 Methodology

Describe the methods, techniques, and tools to be employed. Provide a detailed explanation of your approach. Include algorithms, frameworks, models, and workflows. Use diagrams to explain the methodology clearly.

3.2.1 Overview of the Approach:

A summary of your method.

3.2.2 Dataset Selection:

Describe the datasets (real-world, synthetic, or benchmarks) used for evaluation.

3.2.3 Algorithm/Model Design:

Explain the model architecture, algorithmic steps, or computational processes.

3.2.4 Tools and Technologies:

Mention the programming languages, libraries, or software used.System Architecture/Design (Optional, for implementation-based projects) Provide a high-level view of the system's components and interactions. Include block diagrams or flowcharts to illustrate the design. Describe individual components and their roles.

3.2.5 Algorithm or Model Description

Explain the model architecture, algorithmic steps, or computational processes.

Also give explanation of how these algorithms are used in solving the problems. The prerequisites, if any, need to be explained. Any equations to clarify the algorithms, if necessary, may also be given.

A sample algorithm is shown in Algorithm 1

Algorithm 1 Pseudocode for existing model

1: **Input** Subset of Features with corresponding values; Required Target feature name
2: **Output** Target feature value
3: Retrieve the coefficient values from the database for the given input subset of features
4: **for** each record in the features' coefficient values **do**
5: Apply linear regression
6: $h_\theta(x) \leftarrow \theta^T.X$, where θ and X are column matrices
7: **end for**

Provide detailed steps of the proposed algorithm or architecture. Outline the algorithm step-by-step. Include pseudocode or flowcharts for clarity. Explain any mathematical formulations or equations involved.

3.2.6 Expected Outcomes

Discuss the anticipated results and how they address the research objectives. Highlight the expected performance improvements or practical contributions.

3.2.7 Advantages of the Proposed Work

Emphasize the benefits of the proposed approach. Compare with existing methods, highlighting improvements. Mention scalability, efficiency, accuracy, or usability gains.

3.2.8 Limitations and Assumptions

Acknowledge potential limitations and assumptions made in the study.

Mention constraints like computational resources, data availability, or generalizability. Discuss any assumptions inherent in the model or methodology.

Chapter 4

Experimentation and Result Analysis

Describe how the proposed work was evaluated, and provide an analysis of the results.

4.1 Experimental Setup

Describe the resources, tools, and procedures used for conducting experiments.

Hardware and Software: Specify the computational resources, programming languages, libraries, or platforms used. Example: "Experiments were conducted on a system with an Intel i7 processor, 16GB RAM, using Python 3.9 and PyTorch."

4.2 Datasets:

Provide details about the datasets used, including their source, characteristics (e.g., number of nodes, edges, features), and preprocessing steps. Example: "The experiments used two datasets: the Facebook Social Network dataset and the DBLP Citation Network, both containing time-stamped edges."

4.3 Evaluation Metrics:

Define the metrics used to measure performance, such as precision, recall, F1 score, ROC-AUC, or computational efficiency. Example: "Precision and recall were used to evaluate prediction accuracy, while runtime was used to measure computational efficiency."

4.4 Experimental Design

Describe how the experiments were conducted step by step.: Baseline Methods: List the existing methods used for comparison. Example: "The proposed model was compared against node2vec, GCN, and GAT models."

4.4.1 Experimental Scenarios:

Discuss different conditions under which the experiments were conducted, such as varying dataset sizes, hyperparameters, or network dynamics. Example: "The experiments evaluated the model under static and dynamic conditions by incrementally adding edges over time."

4.4.2 Parameter Tuning:

Mention how hyperparameters were optimized, if applicable. Example: "Grid search was performed to tune the learning rate, dropout rate, and number of GNN layers."

4.5 Results

Present the results of your experiments in a clear and organized manner.

Tables and Graphs: Use tables, charts, and graphs to present quantitative results. Comparative Results: Compare the performance of your proposed model with baseline methods. Example: "The proposed model achieved an F1 score of 0.89 compared to 0.83 for node2vec." Qualitative Results (if applicable): Include visualizations, case studies, or examples of outputs. Example: "Figure 4.2 shows the network before and after link prediction using the proposed model."

4.6 Analysis of Results

Interpret the results and explain their significance.

4.7 Observations:

Discuss patterns or anomalies observed in the results. Example: "The proposed model outperforms baseline methods for dense graphs but shows reduced performance for sparse graphs." Insights: Relate the results back to the research objectives and literature gaps. Example: "The results validate the proposed hybrid approach's effectiveness in handling dynamic networks, addressing the scalability issue highlighted in the literature." Statistical Significance (if applicable): Include significance testing to strengthen claims.

4.8 Comparative Analysis

Provide a side-by-side comparison of your results with those of existing methods.

Include a table summarizing metrics for all methods. Discuss the relative advantages or limitations of your approach. Example: "While the runtime of the proposed model is slightly higher than node2vec, it achieves significantly better precision."

Chapter 5

Conclusion and Scope for further Research

Appendix A

title of appendix

Include your Code here

References

[1] M. Jung, J. S. Song, A.-Y. Shin, B. Choi, S. Go, S.-Y. Kwon, J. Park, S. G. Park, and Y.-M. Kim, "Construction of deep learning-based disease detection model in plants," *Scientific reports*, vol. 13, no. 1, p. 7331, 2023.

Papers Communicated

5. Mr. X, Mr. U, Mr. V and Mr.Z , "Efficient Computations in Operating Systems", (Communicated in Sept. 2018).